Combining Machine Translation and Translation Memory Tools: Implications for English-to-Arabic Technical Translation

Ahlem CHELGHOUM
University of Annaba

Résumé
Recently, translation has witnessed many changes which have given birth to 'translation technology field'. This paper attempts to set aside the classical methods of translation and introduce a new direction that combines both Machine Translation (MT) and Translation Memory (TM) technologies. In the light of an empirical investigation on students' technical translation carried at the Department of English (University of Constantine), we would examine the possibility of improving students' translation quality and speed.

1. Introduction

Translation has been a necessity over years. It is rising more than ever due to globalisation and the need for communication. Besides, translation market has become very competitive where efficiency and speed are fundamentally important element. Thus, technology offers several translation tools, each of which responds to different needs and requirements. Technical translation, for instance, is a very complex process requiring both skills and technological tools for a highly enhanced translation quality and faster processing. This, therefore, has given birth to a new discipline known as 'translation technology' [1]. In its essence, translation technology is mainly the use of computerised or technological tools in order to assist translators in translation [10]. The most commonly known translation technologies are machine translation (MT) and translation memory (TM). Therefore, the ultimate aim of the

present paper is not in line with attempting to replace translators, but rather assist them by combining both technologies for a better and faster technical translation.

2. Machine Translation

Machine Translation, usually referred to by the abbreviation MT, is a sub-field of translation technology that investigates the use of computer software in translation from one natural language, the source language (SL) into the target language (TL) [17][23]. The central core of MT is the automation of the full translation process [26].

2.1 Approaches to Machine Translation

Chéragui (2012) distinguishes two main architectures of MT: linguistic and computational [7].

2.1.1 Linguistic Architecture

There are several approaches that have been made to tackle MT. In the linguistic architecture, there are three approaches that differ in their complexity and sophistication, direct, transfer and Interlingua [2] Fig 1[1].

In the *direct approach* translation is direct from ST to TT; it performs a word-for-word translation [7]. The disadvantage of this approach lies in the poor quality of translation since it rarely uses semantic analysis [22]. In

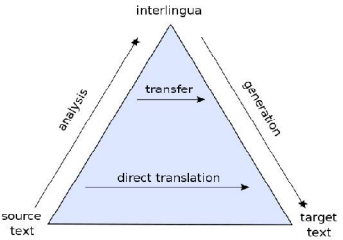

Figure 1 The Pyramid Diagram of Machine Translation Approaches

[1] Vauquois Diagram, 1968 in Somers, 1999, p. 145

the *transfer based machine translation*, the translation is conducted through three stages: the analysis of ST, the transfer into intermediate representation, then the generation of the final TT [7][19]. *The Interlingua approach*, the most suitable for multilingual systems, has two stages; the analysis of sentences with Interlingua representation and the generalization of TT using that representation [2].

2.1.2 Computational Architecture

There are three systems, the first is the *rule-based machine translation (RBMT)*, based on the encoding of linguistic information about SL and TL in the form of rules [28]. The second is the *corpus-based machine translation (CBMT),* also called the data driven approach. It is an alternative approach made to overcome knowledge acquisition which is the RBMT main problem and to generate new translations based on previously stored examples [10][7]. Within the CBMT paradigm, there are two types of MT. *Statistical Machine Translation (SMT)* which is purely statistical, is characterized by the use of machine learning methods via applying algorithm to previously translated texts [20][27]. Flanagan states that SMT is a more dominant approach usually used in commercial translation systems like Google Translate [10]. The *Example-Based Machine Translation (EBMT)* is the second type of MT, situated somehow between RBMT and SMT, since it puts many techniques from both systems together [10]. The three components of EBMT are: first, matching the ST fragments against a database of real examples. Second, identifying the corresponding translation fragments. Finally, recombining these fragments to give the TT [27]. The *Hybrid Approach* is the third computational MT system**.** It combines transfer

approach with one of the corpus-based approaches. It provides some quality improvement through its basic system which consists of learning automatically syntactic transfer rules [7].

2.2 Advantages and Disadvantages of MT

The impressive growth MT has witnessed lastly was due to the notable advantages it has. Flanagan praises MT for being instant, robust and cost-effective. Besides they could be integrated with several tools [10]. Gil and Pym mention the advantages the technology has on all forms of written communication, singling out the MT [13]. MT is not only used for translation purposes, but also for learning. It constitutes a relevant language learning tool for both advanced and beginners mainly in writing and communication [11]. Despite the numerous advantages MT has, shortcomings are inevitable.

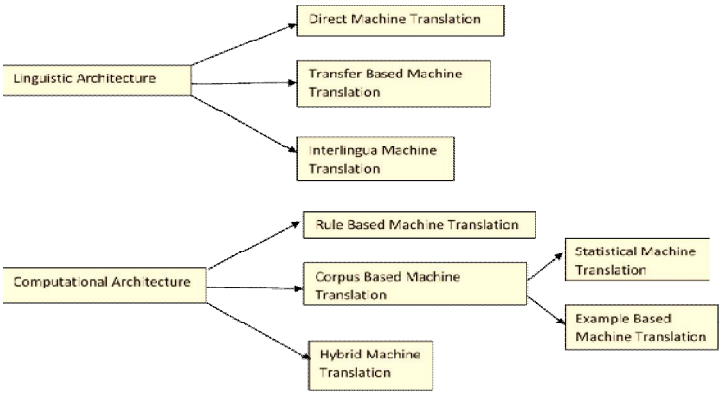

Figure 2 Approaches to Machine Translation (Adapted from Chéragui classification, 2012)

Gross raises three main problems with MT. The first is context and cultural issues that computers are unable to perceive; the second problem is establishing different meanings of humour such as emotions generally found in poetry, and the third is the

inability of MT to deal with ambiguity, idioms and collocations [16]. Arnold, in his turn, distinguishes some practical limitations of MT. He highlights the problems of vaguely specified tasks and the inability of computers to learn new concepts and perform common sense reasoning [3]. In addition to problems with ellipsis, interrogatives, complex sentences, etc. [26].

3. Translation Memory

With globalisation, translation demand has considerably increased and surpassed human translator's capacities. This stimulated researchers to direct their efforts towards MT. This latter which cannot meet all these demands qualitatively, opened the way for research to seek a new technology to provide translators with better assistance and resulted in Translation Memory.

TM systems are programs that create databases of ST and TT segments to be reused [13]. According to Gordon, TM is of a simple concept, as the translator works through a text, segments or TUs will be stored in the database memory and will appear whenever there are similarities in the new segments to be translated [14]. Depending on the degree of similarities, it may be a 100% matching or fuzzy matching. The latter means that the system will find similar segments in the database memory and suggest the previously created translations. Whilst 100% match describes previous TUs stored in the system that perfectly match the new segments [8]. Trujillo further explains that the translator will be able to decide whether to accept the offered translation, modify or reject it. Furthermore, the new translation given by the translator will be automatically saved in the memory [29].

3.1 Differences between MT and TM

The distinction between MT and TM is made by many scholars. Contrary to MT, which is automatic, TM allows the translator to accept or attempt any modification on the suggested segments [22]. Gordon gives a set of differences between MT and TM, where the former is an inflexible system. Also as the outputs are machine production only, the post-editing stage is crucial. Moreover, the MT systems do not represent any learning features. However, TM is a highly flexible system where learning is automatic with an enhanced quality of translation. Moreover, it is fast and precise where all provided outputs are human production [14].

3.2 Approaches to TM

There are two different approaches to TMs, namely the *Sentence-Based (SB)* approach and the *Character-String in Bitext (CSB)* based approach. In the first, the ST and TT are divided into TUs, which can be words, sentences, titles, etc. As the translator works, each created TU is stored in the database for future use, by which the TM system is built [4][5]. The main benefit of SB approach is that sentences are easily identifiable by the computer. Nonetheless, it has problems of 'fuzzy matches' when it creates useless matches called 'noise' or fails in generating useful matches known as 'silence' [4][15]. CSB based approach involves storing of the whole texts with their corresponding translations, known as ***bitexts***. Bitexts can be easily found by the system and reused whatever their length. It tries to look for matches at any string of character [15][21].Yet, this approach increases 'noise phenomenon' that is due to the unreliability of small matches [15].

3.2 Process of TM

TM system uses a set of processes starting from the import of the ST until the post-editing of the TT.

3.2.1 Import and Segmentation

The first stage of translation using a TM tool is to import the text into the chosen system (Transit, TRADOS, Omega T, etc.) [8]. Then comes the segmentation, one of the main features of all TM tools; it is a process of dividing the ST into TUs [22][30].

3.2.2 Alignment and Indexing

After the segmentation comes another process of TM process called 'the alignment'. It consists of creating new TMs, or to add translations to an already existing database [22]. The created TUs in the segmentation will be aligned by matching both source and target segments, then indexed and stored in a terminology database [9][30].

3.2.3 Matching and Retrieval

This process is essential for every TM tool. The system tries to find possible similarities to be proposed to the translator and reused in translation [22][30]. The retrieval is likely to be more efficient depending on the accuracy of both fuzzy and exact matches especially in matter of content [18].

3.2.4 Translation

After the retrieval, the translator can attempt the necessary modifications and keep the useful segments in the TT [8]. In Omega T, for example, the final revision of the TT is carried out in a separate file created in a different folder.

3.3 Advantages of TM Tools

TM systems offer several advantages to translators. Speed and consistency are the most notable ones. For Gow, consistency has a great importance in translating, particularly texts with a higher degree of repetitions like technical texts. Software documentation, online manuals and other digital texts are only subject to some changes or adaptation techniques [13][15]. In addition to the high quality of the outputs and consistency; speed, resulting in time-saving is another significant benefit of TM systems [12] [14]. It is worthy to notice that time-saving in the translation process enables translators to increase their incomes [24].

3.4 Disadvantages of TM

It is highly believed that TMs provide translators with many benefits, but, like any technology, TMs have some limitations. One of the problems posed by using TMs concerns the quality of translation. These tools, for instance, have a negative effect in using cataphoric and anaphoric references [5]. In addition to other obstacles such as the lack of language knowledge and context insensibility, problems of investment such as purchasing the software and the maintenance costs are inevitable [24]. Another problem of TMs is the building of database which is time-consuming. Besides, if the corrections and the modifications attempted in post-editing are done in a separate document, they cannot be captured by the TM system and might be always recycled [22][31][6]. Dennett mentions other problems with regard to using TMs such as segmentation which prevents the translator from dealing with the text as a whole [8]. A possible solution to this problem is training translators to work beyond the segment limits and think about the context of each [6][12]. TMs have also problems of entering ‘foreign language character’, time wasting in post-

editing of the suggested pre-translation ,and the long term required training translators need before using the software [8][22][24].

3.5 Omega T Omega T is a free and open source multiplatform TM tool, written in JAVA and realised under GNU[2] [30]. It is characterised by many features including: TM database, Terminology management (glossaries), segmentation, fuzzy matching, dictionaries, spell checker, alignment tools, translation leveraging into updated projects and many other utilities [30][25]. MT is integrated to Omega T, so that the user can translate non-translated segments. However, translators need to work online in order to use any of the offered MT services including Google Translate, Belazar and Apertium [25].

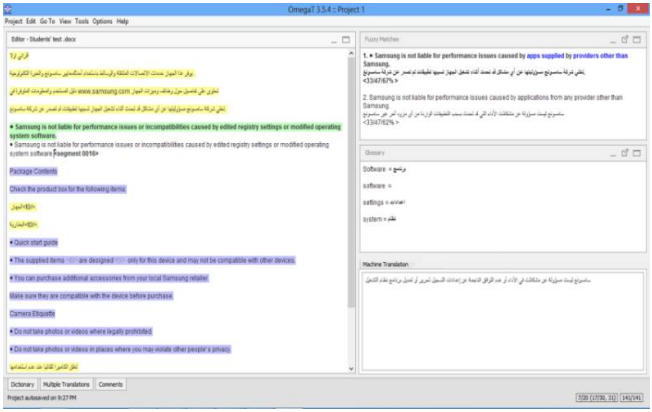

Figure 3: Screen shot of Omega T main window (experimental group participant)

[2] GNU : General Public License

Omega T can use a large number of translation projects. It can support several languages, Unicode (UTF-8), i.e. non-Latin alphabets; it can handle both right to left and left to right languages. Moreover, it can support two types of files, plain text files including ASCII, Portable Object (PO) files, INI files, etc. and formatted text files including MO Open XML, HTML, XHTML, DocBook, etc. [25][30].

4. Experimental Study

The present paper aims to investigate the correlation that may exist between TM and MT on technical translation. It tackles two fundamental features of translation which are quality and speed. An experiment was conducted at the Department of English, University of Constantine, as an attempt to check the validity of stated assumption. To achieve this, a highly representative sample of 26 students was randomly selected from second year master students based on their performance on translation practice. As the main concern of this paper is translation, students were classified into three categories: good (11.11%), average (44.44%), and weak students (44.44%), where the total number of the participants is proportionate to the total number of the population considering their translation marks.

All participants were requested to translate a text of 186 words with a total size of 18 segments from English into Arabic. The test consisted of extracts taken from Samsung mobile phone user manual. The participants were put into two groups, one serves as the control group where students were supposed to translate without assistance. In the experimental group, the subjects were allowed to use both MT and TM systems combined. They were permitted access to their usual reference materials, including digital dictionaries and Internet.

The choice of Google Translate was motivated for being the most popular tool among participants. However, the choice was fallen on Omega T for being free, downloadable and simple tool for beginners, also, for supporting a large number of language pairs including English and Arabic. Note that in order to investigate the effects of using TM tools on the quality of translation, and due to the constraint of time, a TM database was built up with TUs in order to guarantee an appropriate environment for the participants, which fits the purpose of this research paper.

4.1 Discussion of the Results

As previously mentioned, students' results of both groups in the test are evaluated in matter of two parameters: the quality and the speed of translation.

4.1.1 Quality of Translation

The quality of translation is evaluated according to the students' errors in both content and form. For each error, 0.25 is omitted from the mark. According to the obtained results, the overall average of the control group students is 10.17. Concerning the categories, the good students' category obtained an average of 16.5, however, the average and the weak category obtained 10.20 and 9.44 respectively. The percentage of students who got the average in the control group is 53.85% and those who had less than the average is 46.15%. This means that almost half of the students could not provide an acceptable translation. However, the results of the experimental group are very revealing. Students' translation quality has improved significantly. They obtained an average of 18.10. A great improvement in the translation marks is noticed in the three categories. The good, average and weak students' category got respectively an average of 18.25, 18.13 and 18.04. So, one can deduce that students whether good or

weak have improved their translation performance after using MT and TM combined.

After taking a closer look on students' performance on the test; their provided answers were categorised into four categories: correct translation, semi-correct, incorrect and no translation with regard to their frequencies and percentage. In the control group, a low level of correctness was obtained. Most of students provided incorrect translation; only few of them provided semi-correct translation or skipped the translation of segments. The interpretation of their answers shows that most of their mistakes affected the content which reflects their limited vocabulary. In addition, the high percentage of incorrect translations implies that translating technical texts constitutes a major problem for them. Grammatical errors are also widespread such as word order, tenses, punctuation, etc. To illustrate, here are some examples of students' translations:

"This device provides mobile communication and media services using Samsung's latest standards and technological expertise", translated with a 92.31% as frequency of incorrectness as*: " توفر هذه الأدات الاتصالات الهاتفية والخدمات الإعلامية باستعمال مقاييس أسئلة للطلبة ومعدات تكنولوجية "*

"Samsung is not liable for performance issues or incompatibilities caused by edited registry settings or modified operating system software", translated with a 76.92 % as a frequency of incorrectness as:

" سامسونغ ليس مضمون لأداء المشاكل المتداخلة المتسببة من طرف مكان التسجيل أو تصحيح تنظيم العمليات. "

Most students of the experimental group provided correct translations; only few of them gave semi-correct or incorrect translations. When closely analyzing their answers, it was noticed that their problems were with punctuation, word order

or some word choice errors. Nevertheless, the quality of their productions was considerably enhanced. These are some examples of the participants' answers:

"This user manual and the information available at www.samsung.com contain details about the device's functions and features." translated with a 92.31% frequency of correctness as:

« دليل المستخدم هذا والمعلومات المتوفرة على الموقع www.samsung.com يحتويان على تفاصيل حول وظائف وميزات الجهاز »

"The front camera lens is suitable for taking wide-angle photos." translated with 92.31% as a frequency of correctness as:

"الزاوية واسعة صور لالتقاط ملائمة الأمامية الكاميرا عدسة"

4.1.2 Speed of Translation

Speed is an important criterion of a successful translation. In the present paper, students had one hour to complete the test, the time spent in translation was taken by a chronometer. According to the obtained results, the average of time students of the control group spent on translating is 36:02 (min). Most of them spent more than 20 minutes to complete the translation; only in very few instances there were fast students. Whereas the experimental group obtained results denote that students were faster than the control group. They spent an average of 18:29 min to deliver their translation with an even enhanced quality. The post editing stage lasted 02:05 min as an average. Students at this stage were generally correcting some spelling mistakes, using the appropriate punctuation, trying to assure coherence and cohesion. After taking a closer look at how students used Omega T, they were working with a very slow performance at the beginning. This is due to their unfamiliarity with this software. Yet, by the sixth segment, students started to use it smoothly, and managed to speed up

their performance. Therefore, the findings from both groups have enlightened us about students' problems concerning technical translation and the use of technological tools. Considering the obtained results, one can assert that students really lack vocabulary. However, students' familiarity with MT system makes the introduction of a TM not quite difficult. In short, the students have improved their technical translation speed and quality when assisted.

5. Conclusion and Pedagogical Recommendations

The translation field is undergoing vital changes like using software for translation. This creates the idea that these tools may replace the human translators and are considered more as a threat to their profession than assistance. However, this idea cannot be acceptable as the human translator still plays a crucial role in translation. The present research paper aims at investigating the possibility of improving students' technical translation using TM and MT systems combined. Hence, the results were significant. Using these tools showed a great improvement in both quality and time.

In the light of the obtained results, it is recommended to include technological tools in the syllabus of translation practice, where students can have a training course and develop their translation performance and vocabulary. In addition, the integration of TM within MT systems inside and outside classroom is recommended. To sum up, in this digital world, giving more interest to teaching students how to use technology in translation instead of the classical methods of translation is highly advocated.

References

[1] ALCINA, A. 2008. Translation Technologies: Scope, tools and resources. *Target: International Journal of Translation Studies* (33), 79-102.
[2] ANDERSON, D. 2013. Machine Translation as a Tool in Second Language Learning. *CALICO journal, 13*(1), 68-97.
[3] ARNOLD, D. 2003. Why Translation is Difficult for Computers. In H. Somers, *Computers and Translation: A Translator's Guide* (pp.119-142). Amsterdam / Philadilphia: John Benjamins BV.
[4] BOWKER, L. 2002. *Computer-Aided Translation Technology: A Practical Introduction.* Ottawa: University of Ottawa Press.
[5] BRKIĆ, M., SELJAN, S., & MIKULIĆ, B. B. 2009. Using Translation Memory to Speed up the Translation Process. *InFutur: "Digital Resources and Knowledge Sharing»*, 353-363.
[6] CANIM, S. 2010. Translation Memory Systems for Avoiding Context Deficiency. Istanbul, Istanbul University.
[7] CHÉRAGUI 2012. Theoretical Overview of Machine translation. *Proceedings ICWIT*, 160-169.
[8] DENNETT, G. 1995. Translation Memory: Concepts, Products, impact and Prospects. London, South Bank University.
[9] ESSLINK, B. 2000. *A Practical Guide to Localization.* Amesterdam / Philadelphia: John Benjamins Publishing Company.
[10] FLANAGAN, M. 2009. Recycling Texts: Human Evaluation of Example-based Machine Translation Subtitles for DVD. *PhD Thesis*, Dublin City University, Dublin.
[11] GARCIA, I., & PENA, M. I. 2011. Machine Translation-Assisted Language Learning: Writing for Beginners. *Computer Assisted Language Learning, 24*(5), 471-487. doi:10.1080/09588221.2011.582687
[12] GARCIA, I. 2009. Beyond Translation Memory: Computers and the Professional Translator. *The Journal of Specialised Translation, issue 12*, 199-214. Retrieved from http://www.jostrans.org/issue12/art_garcia.php
[13] GIL, J. R., & PYM, A. 2006. Technology and Translation (a Pedagogical 0verview). *Translation Technology and Its Teaching*, 5-19. (A. Pym, A. Perekrestenko, & B. Starink, Eds.) Tarragona: Intercultural Studies. Retrieved May 10, 2007, from http://isg.urv.es/library/papers/BiauPym_Technology.pdf
[14] GORDON, I. 1996. "Letting the CAT out of the Bag- or was it MT? » *Translation and the Computer.18: Papers from the Aslib Conference Held on 14 & 15 November 1996*, Aslib, London.
[15] GOW, F. 2003. Metrics for Evaluating Translation Memory Software. *MA thesis*. University of Ottawa, Canada.

[16] GROSS, A. 1992. Limitations of Computers as Translation Tools. In J. Newton, *Computers in Translation: A practical Appraisal* (pp. 96-130). London: Routledge.
[17] HUTCHINS, J.W. 1986. *Machine Translation: past, present, future (Ellis Horwood Series in Computers and their Applications).* New York: Prentice Hall.
[18] LAGOUDAKI, E. 2006. Translation Memories Survey 2006, Translation Memory Systems: Enlightening user's perspective. In Proceedings of *Aslib International Conference 'Translating and the Computer'28 (pp. 15-16).* London: Aslib.
[19] LAMPERT, A. 2004. *Interlingua in Machine Translation.* Technical Report.
[20] LOPEZ, A. 2008. Statistical Machine Translation. *ACM Computing Surveys, 40* (3 Article 8), pp.1-49.
[21] MCBRIDE, C. 200). Translation Memory System: An Analysis of Translator's Attitudes and Opinions. Ontarion, Ottawa University, Canada.
[22] MIKULIČKOVÁ, E. 2010. Computer Assisted Technology-Comparison of Programs. Zlin, Tomas Bata University: Faculty of Humanities.
[23] Reddy, M., & Hanumanthappa, M. (2013). Indic Language Translation in CLIR Using Virtual Keyboard. In A. Kumar, & R. Selvarni, *Advances in Intelligent Systems and Computing* (p. 174). New Delhi: Springer.
[24] SELJAN, S., & PAVUNA, D. 2006. Translation Memory Database in the Translation Process. *Proceedings of the 17th International Conference on Information and Intelligent Systems IIS.* Varaždin.
[25] SMOLEJ, V. 2013. Omega T 2.6 -User's Guide. From: http://www.omegat.org/fr/documentation.php
[26] SOMERS, L., & HUTCHINS, J. 1992. *An Introduction to Machine Translation.* London, the University of Michigan, United States: Academic Press.
[27] SOMERS, H. 1999. Review Article: Example-based Machine Translation". *Machine Translation, 14,* 113–157. Doi: 10.1023/a: 1008109312730.
[28] SOUDI, A., FARGHALI, A., NEUMANN, G., & ZBIB, R. (Eds.). 2012. *Challenges for Arabic Machine Translation.* Amesterdam /Philadelphia, USA: John Benjamins Publishing Company.
[29] TRUJILLO, A. 1999. *Translation Engines: Techniques for Machine Translation.* (P. R. Kuljis, Ed.) London: Springer-Verlag.
[30] WASALA, R. A. 2007. Initial Survey on the Availability of Translation Memory Tools. *PAN Localization Project(Project N° 102042).* University of Colombo.
[31] WEBB, L. E. 1998. Advantages and Disadvantages of Translation Memory: A Cost/Benefit Analysis. *MA Thesis,* Monterey Institute of International Studies, California.
http://www.webbsnet.com/translation/thesis.html